REMARKS

Claims 1, 3-6, 9, 11-14, 17-27, 31, 35-44, 48, 52, and 53 are pending. Claims 2, 7, 8, 10, 15, 16, 28-30, 32-34, 45-47, and 49-51 have been previously canceled. Claims 4 and 12 are presently canceled. Claims 1, 9, 17, 18, 20, 35, 37, 52, and 53 have been amended. No new matter has been entered. Claims 1, 3, 5, 6, 9, 11, 13, 14, 17-27, 31, 35-44, 48, 52, and 53 remain.

Rejection under 35 U.S.C. § 103(a) over Lindh et al. in view of Dhillon et al.

Claims 1-6, 9-14, 17-23, 35-40, 52, and 53 stand rejected under 35 U.S.C. § 103(a) as being obvious over International Application Publication No. WO 03/060766, to Lindh et al. ("Lindh"), in view of U.S. Patent No. 6,560,597, issued to Dhillon et al. ("Dhillon"). Applicant traverses.

Claims 2 and 10 have been rejected as obvious. However, Claims 2 and 10 were canceled in a Response to Office Action filed on January 25, 2008. Consequently, for purposes of response, the 35 U.S.C. § 103(a) rejection as applied to Lindh and Dhillon is assumed to apply to Claims 1, 3-6, 9, 11-14, 17-23, 35-40, 52, and 53, as specifically discussed *infra*.

Further, the examiner bears the initial burden of factually supporting any *prima facie* conclusion of obviousness, which includes a clear articulation of the reasons or rationale why the claimed invention would have been obvious. MPEP 2142. Exemplary rationales to support a conclusion of obviousness are listed in MPEP 2143, although the list is not all-inclusive.

The claims appear to be rejected under the rationale outlining combining prior art elements according to known methods to yield predicable results, which includes *inter alia* "a finding that the prior art included each element claimed, although not necessarily in a single prior art reference, with the only difference between the claimed invention and the prior art being the lack of actual combination of the elements in a single prior art reference." MPEP 2143(A). If any of the findings cannot be made, this rational cannot be used to support a conclusion that the claim would have been obvious. *Id.* A *prima facie* case of

obviousness has not been shown.

Lindh teaches preprocessing a corpus of documents by performing word splitting, identifying proper names, removing stop words, applying a word stemming algorithm, and performing word weightings (Lindh, p. 19, lines 2-5). Following preprocessing, each unique term is assigned a weight according to that term's information content, which is determined by Term Frequency times Inverse Document Frequency (TFIDF) (Lindh, p. 17, lines 21-23). Matrices are generated to describe relationships within the document corpus using the unique terms (Lindh, p. 18, lines 23-25). A document-concept matrix provides relationships between the documents in the corpus and concepts (Lindh, p. 19, lines 12-21). A term-document matrix provides relationships between the documents and unique terms selected from the documents (Lindh, p. 19, lines 22-28). A term-concept matrix receives information from the document-concept matrix and the term-document matrix to generate weight values representing relationships between the terms and the concepts (Lindh, p. 19, lines 29-32). The term-document matrix and the term-concept matrix are then used to generate a term-term matrix for describing relationships between the unique terms (Lindh, p. 20, lines 22-32). The term-term matrix is used for retrieving information from the document corpus (Abstract).

Lindh further teaches enhancing the above-described relationships by filtering the document corpus (p. 27, lines 18-25). A reduction in the number of similar documents in the corpus precludes large quantities of similar documents from biasing the relationship measures, which is characterized as a flaw that can be reduced using document clustering, such as *k*-means clustering (p. 27, line 25-p. 28, line 5). A representative document vector is generated for each cluster found by a clustering algorithm, such as by calculating a cluster centroid as the mean of all document vectors in the cluster (p. 28, lines 8-23). The representative document vector is added to the cluster and all other documents that belong to the cluster are removed from the initial document corpus (p. 28, lines 8-23).

Independent Claims 1, 9, 17, 18, 35, and 52 have been amended. Claim 1

now recites a scoring module determining a score, which is assigned to at least one concept that has been extracted from a plurality of electronically-stored documents, wherein the score is calculated as a function of a summation of a frequency of occurrence of the at least one concept within at least one such document, a concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight. Claim 9 recites determining a score, which is assigned to at least one concept that has been extracted from a plurality of electronically-stored documents, wherein the score is calculated as a function of a summation of a frequency of occurrence of the at least one concept within at least one such document, a concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight. Claim 17 recites code for determining a score, which is assigned to at least one concept that has been extracted from a plurality of electronically-stored documents, wherein the score is calculated as a function of a summation of a frequency of occurrence of the at least one concept within at least one such document, a concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight.

Similarly, Claim 18 now recites a concept weight module analyzing a concept weight reflecting a specificity of meaning for the at least one concept within the document, wherein the concept weight is based on a number of terms for the at least one concept. Claim 35 recites analyzing a concept weight reflecting a specificity of meaning for the at least one concept within the document, wherein the concept weight is based on a number of terms for the at least one concept. Claim 52 recites code for analyzing a concept weight reflecting a specificity of meaning for the at least one concept within the document, wherein the concept weight is based on a number of terms for the at least one concept. Claim 53 recites means for analyzing a concept weight reflecting a specificity of meaning for the at least one concept within the document, wherein the concept weight is based on a number of terms for the at least one concept. Support for the claim amendments can be found in the specification on page 15, lines 13-28. No

new matter has been entered.

Concept Weight vs. Term Weight.

Lindh teaches determining a term weight for each unique word in a text (Lindh, p. 17, lines 20-23). The term weight is calculated using a TFIDF equation, which includes multiple parameters, including a number of occurrences of a term in a document, a total number of terms in the document, a number of documents in which the term exists, a total number of documents in the document corpus, and a weight function dependent on the positions of the terms in the document (Lindh, p. 17, line 30-p. 18, line 9). The parameters are then entered into the TFIDF equation to calculate the weight for a particular term. Thus, a number of occurrences of a particular terms are determined for a *single document*, as well as a *total number of terms* included in the document. Therefore, Lindh fails to determine a <u>number of terms</u> for <u>a concept</u> as a concept weight.

Score vs. Relation Value.

Lindh also fails to teach a score, which is assigned to at least one concept. Instead, Lindh teaches calculating a relation value for a given term and a given concept (Lindh, p. 22, line 34-p. 23, line 3). The relation value is determined using a given equation, which is based on a term weight and a document-concept relationship value (*Id.*). The term weight is calculated using the TFIDF equation (Lindh, p. 17, line 30-p. 18, line 9) and the document-concept relationship value describes a relationship between a document and a concept (p. 23, lines 7-9). Thus, the relation value fails to consider a concept weight, which is based on a number of terms for a concept. Therefore, Lindh teaches a relation value for a given term and given concept, rather than a score that is calculated as a function of a summation of a frequency of occurrence of at least one concept within at least one such document, a concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight, per Claims 1, 9, 17, 18, 35, 52, and 53.

Score vs. Relationship Value.

Lindh also teaches a method for finding biased information to allow a user

to identify concepts of particular interest (Lindh, p. 30, lines 16-18). A concept bias engine retrieves a set of relevant documents that are related to at least one term and at least one concept, which is provided by a user or search engine (Lindh, p. 29, lines 22-30). Once provided, the concept will "bias" the set of relevant documents to be selected (Lindh, p. 29, line 31-p. 30, line 4). If no concept is provided, the retrieved documents are related only to the term without any bias (Lindh, p. 29, lines 31-34). A method for finding the biased information includes finding documents that contain a given term (Lindh, p. 30, lines 16-31). Concept distributions associated with the document are identified. A user selects one or more of the associated concepts, which creates an input bias conceptual distribution (*Id.*). A relationship value is calculated for each document according to a given equation in which weights for the conceptual distribution and the input bias conceptual distribution are summed over every concept (*Id.*). Lindh fails to provide *how* the conceptual distribution weights are determined. In addition, the relationship value for each document fails to consider a frequency of occurrence of at least one concept within at least one such document, a concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight. Therefore, Lindh teaches determining a relationship value for identifying documents that are biased by a user selection, rather than determining a score per Claims 1, 9, 17, 18, 35, 52, and 53.

Similarity as an Inner Product vs. Relationship Value

Amended Claim 1 further recites forming the score assigned to the at least one concept as a normalized score vector for each such document and determining a similarity between the normalized score vector for each such document as an inner product of each normalized score vector. Claims 9, 17, 18, 35, 52, and 53 recite limitations consistent with Claim 1, as amended.

Lindh fails to teach such limitations. Instead, Lindh teaches allowing a user to select one or more search terms for which related concepts are returned (Lindh, p. 30, lines 7-10). The user then selects one or more of the related concepts and documents that concern both the search term and the selected

concepts are returned (Lindh, p. 30, lines 10-15). The concept selected by a user introduces bias to the set of documents and rearranges the set (Lindh, p. 29, line 31-p. 30, line 2). To locate the biased information, a document corpus is generated based on selected terms (Lindh, p. 30, lines 18-21). A relationship value for each document is calculated from a document conceptual distribution and an input bias conceptual distribution received from a user (Lindh, p. 30, lines 23-30). A sum of the document conceptual distribution and the input bias conceptual distribution is calculated over every concept (*Id.*). Documents that are related to both the document conceptual distribution and input bias conceptual distribution are returned (Lindh, p. 30, lines 10-15; p. 30, line 31-p. 31, line 3). The document conceptual distribution and the input bias conceptual distribution are considered over all concepts to identify biased information, instead of comparing similarity values for each document. Thus, Lindh teaches returning documents based on a relationship value that includes input bias conceptual distributions, rather than determining a similarity between a normalized score vector for <u>each document</u> as an <u>inner product</u> of each normalized score vector.

Selecting Candidate Seed Documents vs. Applying a Cluster Algorithm

Amended Claim 1 further recites a selection submodule selecting a set of candidate seed documents selected from the plurality of documents, a seed document identification submodule identifying a set of seed documents by applying the similarity to each such candidate seed document and selecting those candidate seed documents that are sufficiently unique from other candidate seed documents as the seed documents, a non-seed document identification submodule identifying a plurality of non-seed documents, a comparison submodule determining the similarity between each non-seed document and a center of each cluster, and a clustering submodule grouping each such non-seed document into a cluster with a best fit, subject to a minimum fit. Claims 9, 17, 18, 35, 52, and 53 recite limitations consistent with Claim 1, as amended.

In contrast, Lindh teaches document clustering to reduce a number of similar documents in a document corpus to prevent relationship bias between

terms (Lindh, p. 27, lines 25-28). Clusters are identified by a clustering algorithm, such as a k-means algorithm (Lindh, p. 28, lines 3-11). A representative document vector, generated by the clustering algorithm for each cluster identified, is determined by calculating a cluster centroid as the mean of all document vectors in the cluster (Lindh, p. 28, lines 11-14). The calculated representative document vector is then added to the cluster (Lindh, p. 28, lines 14-16). After determining a representative document vector for each cluster, a new document corpus is produced, in which each cluster is represented by a cluster representative vector (Lindh, p. 28, lines 20-23). The clustering algorithm is applied to the complete document corpus (Lindh, p. 28, lines 9-11), instead of being applied to a select portion of the document corpus. Thus, Lindh teaches applying a clustering algorithm to a document corpus, rather than selecting a set of candidate seed documents from a plurality of documents.

Identifying a Set of Seed Documents vs. Applying a Cluster Algorithm

Further, Lindh fails to teach identifying a set of seed documents from the set of candidate seed documents. As described above, Lindh applies a clustering algorithm, such as a k-means algorithm to a document corpus to remove documents that are similar. After the clusters have been identified, a representative document vector is determined and assigned to each cluster (Lindh, p. 28, lines 20-23). Next, the representative document vector is added to the cluster, and documents belonging to the cluster are removed except for the representative document vector (Lindh, p. 28, lines 16-24; FIGURE 9A). As the clustering algorithm is applied to the complete document corpus, a set of candidate seed documents are not selected, nor is a set of seed documents identified based on a similarity determined for each document. Thus, Lindh teaches applying a clustering algorithm to a document corpus to identify clusters of the documents, rather than identifying a set of seed documents by applying the similarity to each such candidate seed document in each category and selecting those candidate seed documents that are sufficiently unique as the seed documents.

Grouping Non-Seed Documents vs. Applying a Clustering Algorithm

Moreover, Lindh fails to teach assigning non-seed documents into a cluster with a best fit, subject to a minimum fit. Instead, Lindh teaches a clustering algorithm, such as *k*-means clustering (Lindh, p. 28, lines 9-11). A set of clusters containing similar documents will be produced (Lindh, p. 28, lines 6-7). Thus, each document will be clustered with similar documents based on a particular algorithm without applying further requirements, such as a minimum fit criterion. Applying a minimum fit criterion to the teachings of Lindh would change the clustering of the documents since each document must satisfy additional criteria. For example, a document that is similar to a cluster will be placed in that cluster according to the clustering algorithm in Lindh. However, if minimum criteria were applied, that same document may not be placed into the cluster, even though the cluster is similar, if the similarity fails to meet a minimum similarity. Therefore, Lindh teaches assigning documents to similar clusters using a clustering algorithm, rather than grouping a non-seed document into a cluster with a best fit, <u>subject to a minimum fit</u>.

Accordingly, a *prima facie* case of obviousness has not been shown with respect to independent Claims 1, 9, 17, 18, 35, 52, and 53. A similar conclusion would adhere under the other exemplary rationales in the *KSR* Guidelines to fail to demonstrate obviousness. MPEP 2143.

Dependent Claims

Further, Lindh fails to teach the limitations of dependent Claims 19 and 36. Claim 19 recites the scoring module evaluating the score in accordance with the formula:

$$S_i = \sum_{1 \to n}^{j} f_{ij} \times cw_{ij} \times sw_{ij} \times rw_{ij}$$

where S_i comprises the score, f_{ij} comprises the frequency, $0 < cw_{ij} \leq 1$ comprises the concept weight, $0 < sw_{ij} \leq 1$ comprises the structural weight, and $0 < rw_{ij} \leq 1$ comprises the corpus weight for occurrence j of concept i. Dependent Claim recites limitations consistent with Claim 19.

Instead, Lindh teaches calculating a relation value for a given term and a given concept (Lindh, p. 22, line 34-p. 23, line 3). The relation value is determined using a given equation, which is based on a term weight and a document-concept relationship value (*Id.*). The term weight is calculated using a TFIDF equation (Lindh, p. 17, line 30-p. 18, line 9) and the document-concept relationship value describes a relationship between a document and a concept (p. 23, lines 7-9). Thus, the relation value fails to consider a concept weight, which is based on a number of terms for a concept. Therefore, Lindh teaches a relation value for a given term and given concept, rather than a score that is calculated according to the <u>equation</u> of dependent Claims 19 and 36.

Lindh also teaches a method for finding biased information to allow a user to identify concepts of particular interest (Lindh, p. 30, lines 16-18). A concept bias engine retrieves a set of relevant documents that are related to at least one term and at least one concept, which is provided by a user or search engine (Lindh, p. 29, lines 22-30). Once provided, the concept will "bias" the set of relevant documents to be selected (Lindh, p. 29, line 31-p. 30, line 4). If no concept is provided, the retrieved documents are related only to the term without any bias (Lindh, p. 29, lines 31-34). A method for finding the biased information includes finding documents that contain a given term (Lindh, p. 30, lines 16-31). Concept distributions associated with the document are identified. A user selects one or more of the associated concepts, which creates an input bias conceptual distribution (*Id.*). A relationship value is calculated for each document according to a given equation in which weights for the conceptual distribution and the input bias conceptual distribution are summed over every concept (*Id.*). Lindh fails to provide how the concept weights are determined. In addition, the relationship value for each document fails to consider a frequency of occurrence of at least one concept within at least one such document, concept weight based on a number of terms for the at least one concept, a structural weight, and a corpus weight. Therefore, Lindh teaches determining a relationship value for identifying documents that are biased by a user selection, rather than determining a score

according to the equation of dependent Claims 19 and 36.

Furthermore, dependent Claim 20 recites the concept weight module evaluating the concept weight in accordance with the formula:

$$cw_{ij} = \begin{cases} 0.25 + \left(0.25 \times t_{ij}\right), & 1 \le t_{ij} \le 3 \\ 0.25 + \left(0.25 \times \left[7 - t_{ij}\right]\right), & 4 \le t_{ij} \le 6 \\ 0.25, & t_{ij} \ge 7 \end{cases}$$

where cw_{ij} comprises the concept weight and t_{ij} comprises a number of terms for occurrence j of each such concept i. Dependent Claim 37 recites limitations consistent with Claim 20.

In contrast, Lindh teaches determining a term weight for each unique word in a text (Lindh, p. 17, lines 20-23). The term weight is calculated using a TFIDF equation, which includes multiple parameters, including a number of occurrences of a term in a document, a total number of terms in the document, a number of documents in which the term exists, a total number of documents in the document corpus, and a weight function dependent on the positions of the terms in the document (Lindh, p. 17, line 30-p. 18, line 9). The parameters are then entered into the TFIDF equation to calculate the weight for a particular term. Thus, a number of occurrences of a particular term is determined for a *single document*, as well as a *total number of terms* included in the document. Therefore, Lindh fails to calculate a concept weight, based on a number of terms for a concept, according to the equation of Claims 17 and 37.

Moreover, Claims 3, 5, and 6 are dependent on Claim 1 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 11, 13, and14 are dependent on Claim 9 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 19-27 and 31 are dependent on Claim 18 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Claims 36-44 and 48 are dependent on Claim 35 and are patentable for the above-stated reasons, and as further distinguished by the limitations therein. Withdrawal of the

rejection is requested.

Rejection under 35 U.S.C. § 103(a) over Lindh and Dhillon et al. as applied to Claims 18 and 35, and further in view of Lin et al.

Claims 24-27 and 41-44 stand rejected under 35 U.S.C. § 103(a) as being obvious over Lindh and Dhillon as applied to Claims 18 and 35 above, and further in view of U.S. Patent No. 6,675,159, issued to Lin et al. ("Lin"). Applicant traverses.

Adding the teachings of Lin to the teachings of Lindh and Dhillon introduces further functionality. However, as discussed above, Lindh and Dhillon fail to render Claims 18 and 35 obvious, and the addition of Lin does no more to support an obviousness rejection of dependent Claims 24-27 and 41-44. Claims 24-27 are dependent upon Claim 18 and are patentable for the reasons stated above, and as further distinguished by the limitations therein. Claims 41-44 are dependent upon Claim 35 and are patentable for the reasons stated above, and as further distinguished by the limitations therein. Withdrawal of the rejection is requested.

Rejection under 35 U.S.C. § 103(a) over Lindh and Dhillon et al. and further in view of Lin et al.

Claims 30, 31, 47, and 48 stand rejected under 35 U.S.C. § 103(a) as being obvious over Lindh and Dhillon, and further in view of Lin. Applicant traverses.

Claims 30 and 47 have been previously canceled. Also, adding the teachings of Lin to the teachings of Lindh and Dhillon introduces further functionality. However, as discussed above, Lindh and Dhillon fail to render Claims 18 and 35 obvious, and the addition of Lin does no more to support an obviousness rejection of dependent Claims 31 and 48. Claim 31 is dependent upon Claim 18 and is patentable for the reasons stated above, and as further distinguished by the limitations therein. Claim 48 is dependent upon Claim 35 and is patentable for the reasons stated above, and as further distinguished by the limitations therein. Withdrawal of the rejection is requested.

The prior art made of record and not relied upon has been reviewed by the applicant and is considered to be no more pertinent than the prior art references already applied.

Claims 1, 3, 5, 6, 9, 11, 13, 14, 17-27, 31, 35-44, 48, 52, and 53 are believed to be in condition for allowance. Entry of the foregoing amendments is requested and a Notice of Allowance is earnestly solicited. Please contact the undersigned at (206) 381-3900 regarding any questions or concerns associated with the present matter.

Respectfully submitted,

Dated: September 2, 2008

By: ______________________
Krista A. Wittman, Esq.
Reg. No. 59,594

Cascadia Intellectual Property
500 Union Street, Ste 1005
Seattle, WA 98101

Telephone: (206) 381-3900
Facsimile: (206) 381-3999

OA 5 Resp